REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*
28-09-2004

2. REPORT TYPE
Final Report

3. DATES COVERED *(From – To)*
15 September 2003 - 24-Nov-04

4. TITLE AND SUBTITLE
Cellular Neural Network (CNN)

5a. CONTRACT NUMBER
FA8655-03-1-3047

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Csaba Rekeczky

5d. PROJECT NUMBER

5d. TASK NUMBER

5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AnaLogic Computers Ltd.
13-17 Kende Street
Budapest H-1111
Hungary

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
EOARD
PSC 802 BOX 14
FPO 09499-0014

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
SPC 03-3047

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report results from a contract tasking AnaLogic Computers Ltd. as follows: The Grantee will investigate a real-time multi-target tracking and actuating system. The tracking algorithm will be implemented on the Cellular Neural Network (CNN) visual microprocessor technology (called the 'ACE-BOX Visual Computer' with the ACE16K CNN chip). Tracking will include selective identification based on shape and size. A high-speed CMOS camera will provide image streams of moving objects displayed on a screen to the Visual Computer. The system will be connected to a switchable laser actuator (such as a laser pointer) which will be used to mark the targets. The proposed schematic diagram of the system and other details are provided in the technical proposal and amendment.

15. SUBJECT TERMS
EOARD, Cellular Neural/Nonlinear Networks, Tracking, Visual Microprocessor

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18, NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			/signed/PAUL LOSIEWICZ, Ph. D.
UNCLAS	UNCLAS	UNCLAS	UL		**19b. TELEPHONE NUMBER** *(Include area code)* +44 20 7514 4474

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39-18

Contract number: FA8655-03-1-3047
Expiration date: August 31, 2004
Total dollar value: $25, 000

SENSING-COMPUTING-ACTUATING MULTI TARGET TRACKING SYSTEM

Principal Investigator: Csaba Rekeczky, Ph.D.
Technical Manager and Correspondent: Gergely Tímár, M.Sc.
e-mail: timar@eutecus.com, rcsaba@analogic-computers.com
tel: (+361)-279-6256, fax: (+361)-279-6250

Final Report

August 12, 2004

Prepared by AnaLogic Computers Inc
For
EOARD

The Contractor, AnaLogic Computers Ltd., hereby agrees to the following FAR clause:

252.235-7010 ACKNOWLEDGEMENT OF SUPPORT AND DISCLAIMER (MAY 1995)

(a) The Contractor shall include an acknowledgment of the Government's support in the publication of any material based on or developed under this contract, stated in the following terms: This material is based upon work supported by the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory, under Contract No. FA8655-03-1-3047

(b) All material, except scientific articles or papers published in scientific journals, must, in addition to any notices or disclaimers by the Contractor, also contain the following disclaimer: Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory.

DATE: ______________________________

Name and Title of Authorized Official: ______________________________
Ákos Zarándy PhD
CEO of AnaLogic

(End of clause)

DISCLAIMER:

The Contractor, AnaLogic Computers Ltd., hereby declares that, to the best of its knowledge and belief, the technical data delivered herewith under Contract No. FA8655-03-1-3047 is complete, accurate, and complies with all requirements of the contract.

DATE: ______________________________

Name and Title of Authorized Official: ______________________________
Ákos Zarándy PhD
CEO of AnaLogic

(End of Clause)

I certify that there were no subject inventions to declare as defined in FAR 52.227-13, during the performance of this contract.

DATE: _________________________

Name and Title of Authorized Official: ______________________________
Ákos Zarándy PhD
CEO of AnaLogic

SENSING-COMPUTING-ACTUATING MULTI TARGET TRACKING SYSTEM

AnaLogic Computers Inc.

Introduction and Objectives

The aim of the current project was to produce a system capable of tracking and visually tagging 6-8 targets maneuvering rapidly in a rectangular area at frame rates of up to 60 frames per second. To achieve this goal, the proposed system utilizes two different processors: a CNN-based mixed-signal image processor and a digital signal processor (DSP). Input is provided to the system from a high-speed CMOS imager and the targets are tagged by a laser deflector unit.

We devised a simplified experimental setup to help us develop the algorithms and verify their behavior. In this setup, the targets are generated by a separate computer and displayed by a projector onto a screen. This has two advantages: all of the correct target positions are known so there is a baseline truth to which we can compare the output of the tracking algorithms. At the same time, the projector is capable of projecting targets very rapidly (up to 100 frames/sec) thus providing a way for us to test the speed of the tracking system in a controllable manner.

We started development of the image processing algorithms on the Ace4k CNN-UM processor [4], because the software environment (programming SDK) for the Ace16k was still under development and we have not had access to sufficient number of chips. Chip supply problems have been resolved and the system now utilizes the Ace16k chip for critical image processing tasks. The main difference between the two processors is that while the Ace16k has 128x128 cells [6], the Ace4k has only 64x64. The target tracking algorithms and laser control are run on the DSP adjacent to the Ace16k chip.

This report describes the algorithmic structure and the experimental results obtained with the final system.

System architecture

Figure 1 shows the main building units of the MTT system. The input image is acquired by a high-speed CMOS camera capable of capturing 128x128-sized images at 500 frames/sec given sufficient illumination. This input is captured by an industrial PC that also houses the ACE-BOX visual computer. This contains the Ace4k or Ace16k processor, a Texas Instruments TMS320C6202 digital signal processor and 16 MB of RAM. It communicates with the host PC via a 33Mhz PCI bus interface. The CNN-UM chips are responsible for the image processing tasks and part of the feature extraction. After image acquisition, they perform image enhancement to compensate for ambient lighting changes, motion extraction, related image processing tasks and feature extraction for some types of features. The DSP runs the rest of the feature extraction routines, and the motion

correspondence algorithms such as distance calculation, gating, data assignment and target state estimation. The laser controller uses these target parameters to move the laser to the correct positions and illuminate the targets.

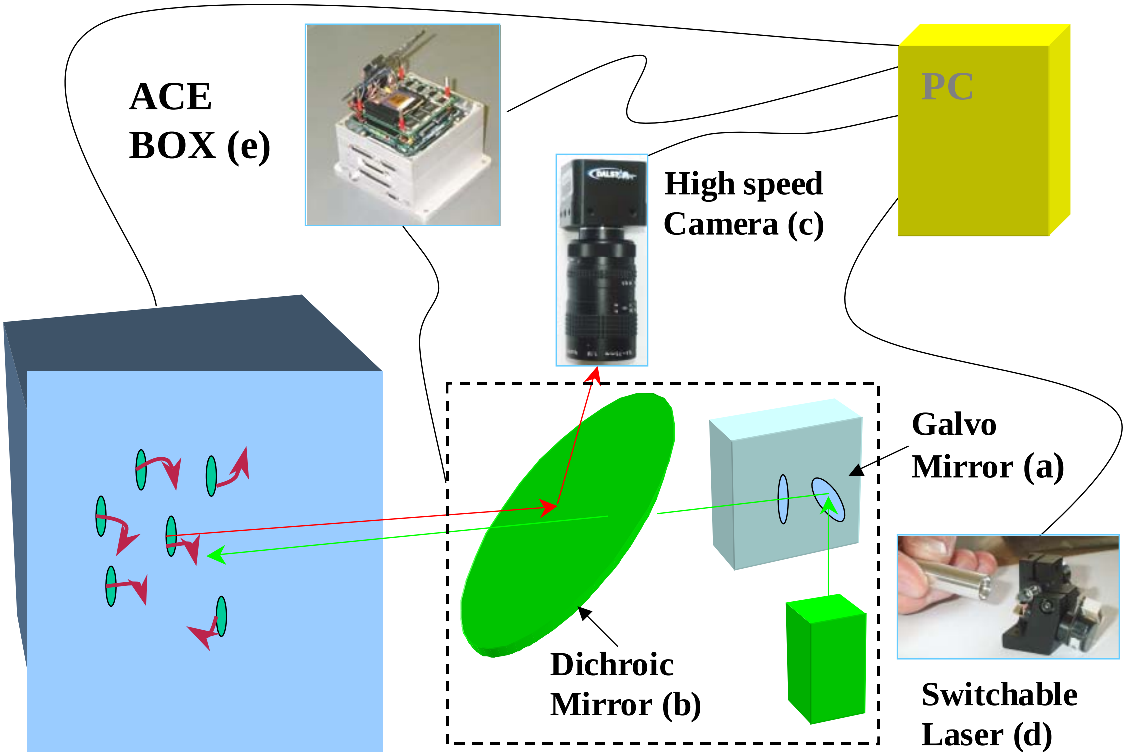

Figure 1. Block diagram of the proposed MTT system

The CNN-UM Algorithms

We tried to capture the main ideas from the natural system by defining three "change enhancing" channels on the input image flow: a spatial, a temporal and a spatio-temporal channel. The spatial channel contains the response of filters that detect spatial i.e. brightness changes, revealing the edges in a frame. The temporal channel contains the result of computing the difference between two consecutive frames, thereby giving a response to motion, while the spatio-temporal channel contains the non-linear combination of the spatial and temporal filter responses. In a general scheme, it can also be assumed that the input flow is preprocessed (enhanced) by a noise suppressing reconstruction filter.

The change enhancement on the parallel channels can be defined as causal recursive difference-type filtering using some linear or nonlinear filters as prototypes (e.g. DoG: difference of Gaussian filtered images implemented by constrained linear diffusion, or DoM: difference of morphology filtered images implemented by min-max statistical filters [5]). It is important to note that in all of these approaches the change enhancing filter channels can be described only by a spatial scale, temporal scale and an orientation parameter. The output of these channels is filtered through a sigmoid type characteristic specified by a threshold and slope parameter.

The output of the best performing individual channel could be used by itself as the output of the image processing front-end, if the conditions where the system is deployed are static and well controlled. If the conditions are dynamic or unknown a priori, then there is no way to predict the best performing channel in advance. Furthermore, even after the system is running, no automatic direct measurement of channel performance can be given short of a human observer deciding which output is the best. To circumvent this problem, we decided to combine the output of the individual channels through a so-called interaction matrix, and use the combined output for further processing. Our experimental results and measurements indicate that the combined output is on average more accurate, than each single channel for different image sequences. Figure 2 shows the conceptual block diagram of the multi-channel spatio-temporal algorithm with all computing blocks to be discussed in the following section.**

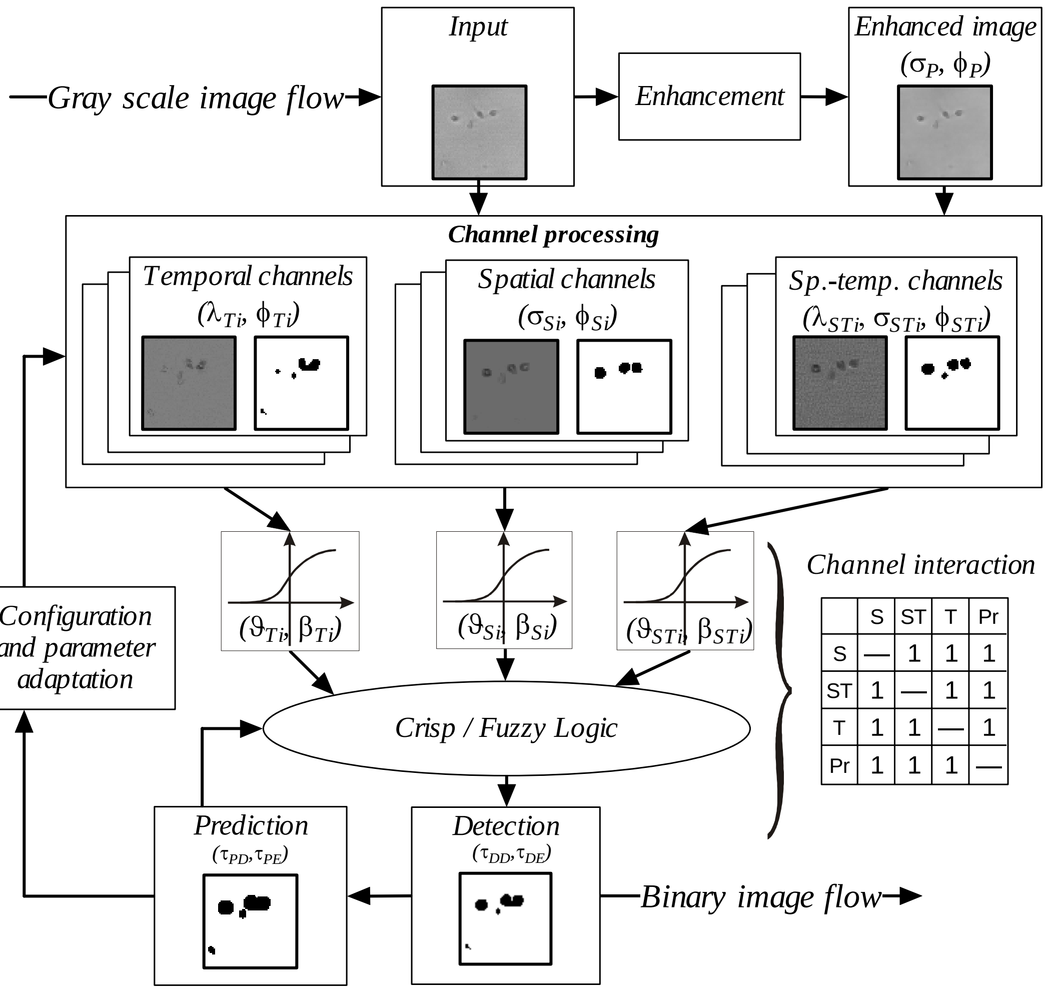

Figure 2 Block overview of the channel-based image processing algorithm for motion detection

The change enhancing channels are actually computed serially (time multiplexed) in any current implementation, but this is not a problem due to the high speed of the CNN-UM chips used. The output of all three channels is a grayscale image that may be

thresholded or processed through a non-linear sigmoid type function. In the first stage of the on-going experiments, only isotropic spatio-temporal processing has been considered followed by crisp thresholding through a hard nonlinearity. Thus, the three types of general parameters used to derive and control the associated CNN templates (or algorithmic blocks) are the scale parameters and the threshold parameter. The enhancement (smoothing) techniques have been implemented in the form of nearest neighbor convolution filters (circular positive *B* template with entries normalized to 1) and applied to the actual frame.

The spatio-temporal channel filtering (including the temporal filtering solution) has been implemented as a fading memory nearest neighbor convolution filter applied to the actual and previous frames. In temporal filtering configuration (no spatial smoothing), λ represents the fading rate (in temporal steps), thereby specifying the temporal scale of the difference enhancement. In the spatio-temporal filtering configuration (the fading rate is set to a fixed value), the scale parameter represents the spatial scale (in pixels) at which the changes are to be enhanced (the number of convolution operations on the current and the previous frame are calculated implicitly from this information).

The pure spatial filtering is based on Sobel-type spatial processing of the actual frame along horizontal-vertical directions and combining the outputs into a single "isotropic" solution.

The change enhancing channels are actually computed serially in the current implementation, but this is not a problem due to the high speed of the CNN-UM chips used. The output of all three channels is a grayscale image that may be thresholded or processed through a non-linear sigmoid type function.

Channel Interaction and Detection Strategies

The interaction between the channels may be Boolean logic based for binary images or fuzzy logic based for grayscale images, specified via the so-called channel interaction matrix.

The interaction matrix is a square matrix where each row and column stands for a single channel in addition to the detection and prediction maps. A row-wise (R) and a column-wise (C) operator must be given that specifies the functions to be used within the rows and between the results. If a cell contains 1 then the given map in the given column must included as is, if it is 0, then it should not be included and if it is –1, then it should be inverted. The interaction matrix allows us to specify very different relationships between the channels within the same framework. For the R and C functions, meaningful spatial logic functions can be selected (e.g. AND – "excitation", XOR – "suppression", OR "summation") resulting in the final output. We found during the experiments that setting R to AND and C to OR works well in most cases.

The result of the channel interaction is a binary map called the detection map that will be the basis for further processing. Ideally, this contains only black blobs where the moving targets are located.

Prediction Methods

We also compute a prediction map that specifies the likely location of the targets in the image solely based on the current detection map and the previous prediction. This can then

be used (via the interaction matrix) as a mask to filter out spurious signals. It is extremely hard to include any kind of kinematical assumption at the cellular level of processing given the real-time constraints, since this would require the generation of a binary image based on the measurements, the current detection and the kinematical state parameters. Therefore, the algorithms only use isotropic maximum displacement estimation implemented by spatial logic and trigger-wave computing. However, the experiments indicate that even rudimentary input masking can be very helpful in obtaining better MTT results.

Figure 3 shows sample frames and their processed output from a test video. This sequence contains 68 frames of seagulls moving rapidly in front of a cluttered background. The black blobs show the birds detected by the multi-channel image processing front-end. This input is used by the feature extractors to determine target positions.

Figure 3 Sample frames from the "birds" test video and corresponding frames from the detection output of the system. Moving targets are circled on the original video

Feature Extraction and Target Filtering

The DSP state-estimation and data assignment algorithms operate on position measurements of the detected targets, therefore these have to be extracted from the detection map. During data extraction, it is also possible to filter targets according to certain criteria based on easily (i.e. rapidly) obtainable features. The set of features we are currently using are: area, centroid, bounding box, equivalent diameter (diameter of a circle with same area), extent (the proportion of pixels in the bounding box that are also in the object), major and minor axis length (the length of the major axis of the ellipse that has the same second-moments as the object), eccentricity (eccentricity of the ellipse that has the same second-moments as the object), orientation (the angle between the *x*-axis and the major axis of the ellipse that has the same second-moments as the object) and the extremal points. Filtering makes possible to concentrate on only a certain class of targets while ignoring others.

The calculation of all of these features can be implemented on the DSP but some of the features (centroid, horizontal or vertical CCD etc.) can be efficiently computed on the CNN-UM as well. Since the detection map is already present on the CNN-UM, calculation

of these features can be extremely fast. It is also possible to calculate a set of features in parallel on the DSP and the CNN-UM, speeding up this processing step even further. The location of the center of gravity (centroid) of each target is usually considered the position of the target, unless special circumstances dictate otherwise.

The DSP-based MTT algorithms

The combined estimation and data association problem of MTT has traditionally been one of the most difficult problems to solve. To describe these algorithms, we need to define some terms and symbols. A *track* is a state trajectory estimated from the observations (measurements) that have been associated with the same target. *Gating* is a pruning technique to filter out highly unlikely candidate associations. A *track gate* is a region in measurement space in which the true measurement of interest will lie accounting for all uncertainties with a given high probability [8]. All measurements within the gating region are considered candidates for the data association problem. Once the existence of a track has been verified, its attributes such as velocity, future predicted positions and target classification characteristics can be established. The *tracking function* consists of the estimation of the current state of the target based on the proper selection of uncertain measurements and the calculation of the accuracy and credibility of the state estimate. Degrading this estimate are the model uncertainties due to target maneuvers and random perturbations, and measurement uncertainties due to sensor noise, occlusions, clutter and false alarms.

Data association

Data association is the linking of measurements to the measurement origin such that each measurement is associated with at most one origin. For a set of measurements and tracks each measurement/track pair must be compared to decide if measurement *i* is related to track *j*. For *m* measurements and *n* tracks, this means *m*n* comparisons, and for each comparison multiple hypotheses may be made. As *n* and *m* increase in number, the problem becomes computationally very intensive. Additionally, if the sensors are in an environment with significant noise and many targets, then the association becomes very ambiguous.

There are two different approaches to solving the data association problem: (i) deterministic (assignment) – the best of several candidate associations is chosen based on a scoring function (accepting the possibility that this might not be correct) (ii) probabilistic (Bayesian) association – use classical hypothesis testing (Bayes' rule), accepting the association hypothesis according to a probability of error, but treating the hypothesis as if it were certain.

Based on data in the literature [8], we decided to work with assignment algorithms because they are high performance with calculable worst case performance since they have a computational complexity of $O(n^3)$ (where n is the number of tracks and measurements) which was essential given our real-time constraints. We also restricted ourselves to the so-called 2-D assignment problems where the assignment depends only on the current and previous measurements (frames). The data assignment algorithms perform so-called *unique* assignment, where each measurement is assigned to one-and-only-one track as

opposed to ***non-unique*** **assignment, when a measurement may belong to multiple tracks. We implemented two types of assignment algorithms a NN approach and the JVC algorithm. Since non-unique assignment would very useful in certain situations such as occlusions, we modified the NN algorithm and added a non-unique assignment mode to it.**

2-D Assignment algorithms

The NN algorithm is the faster algorithm and for situations without clutter works adequately. It can be run in unique assignment mode, where each track is assigned one and only one measurement (the one closest to it) and in non-unique assignment mode, when all measurements within a track's gate are assigned to the track which makes it possible handle cases of occlusion.

The JVC algorithm is implemented as described in [7]. It seeks to find a unique one-to-one track to measurement pairing as the solution $\hat{x}_{ij}$ **to the following optimization problem:**

$$min\left(\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \right) \tag{1}$$

$$\sum_{i=1}^{n} x_{ij} = 1, \sum_{j=1}^{n} x_{ij} = 1 \tag{2}$$

$$0 \leq x_{ij} \leq 1 \quad \forall i, j \tag{3}$$

where n **is the number of tracks and measurements (it is easy to generalize the algorithm if there are more measurements than tracks),** $i,j=1 \ldots n$, c_{ij} **is the probable cost of associating measurement** i **with track** j **calculated based on the distance between the track and the measurement and** x_{ij} **is a binary assignment variable such that**

$$x_{ij} = \begin{cases} 1 & if \ \ j \ is \ assigned \ to \ i \\ 0 & otherwise \end{cases} \tag{4}$$

The JVC algorithm consists of two steps, an auction-algorithm-like step [9] followed by a modified version of the Munkres algorithm [10] for sparse matrices.

Our experiments indicate that the JVC algorithm is indeed superior to the nearest neighbor strategy while only affecting the execution time marginally.

Track Maintenance

We have devised a state machine for each track for easier management of a track's state during its lifetime. Each track starts out in the 'Free' state. If there are unassigned measurements after an assignment run, the remaining measurements are assigned to the available 'Free' tracks and they are moved to the 'Initialized' state. If in the next frame the 'Initialized' tracks are assigned measurements, they become 'Confirmed'; otherwise, they are deleted and reset to 'Free'. If a 'Confirmed' track is not assigned any measurement in a frame, the track becomes 'Unconfirmed'. If in the next frame it still doesn't get a measurement, it becomes 'Free', i.e. the track is deleted.

State Estimation

In this first stage of the experiments, we assume that the target state evolves according to a known linear direct discrete time-varying model. The associated dynamic equations are:

$$x(k+1)=F(k)x(k)+G(k)u(k)+\Gamma(k)v(k);\ k=0,1,... \qquad \textbf{(5)}$$

$$z(k)=H(k)x(k)+\omega(k);\ k=1,... \qquad \textbf{(6)}$$

$$E\left[v(k)v(k)'\right]=Q(k);\ E\left[\omega(k)\omega(k)'\right]=R(k) \qquad \textbf{(7)}$$

where $x(k)$ is the n_x-dimensional state vector, $u(k)$ is an n_u-dimensional known input vector (control or sensor platform motion), while $v(k)$ and $\omega(k)$ are the uncorrelated zero-mean white Gaussian *process noise* and *measurement noise*, respectively (linear Gaussian assumption: $Q(k)$ and $R(k)$ are the corresponding covariance matrices). In the formulation above $F(k)$ is the state transition matrix, $G(k)$ is the input gain, $\Gamma(k)$ is the process noise gain and $H(k)$ is the measurement matrix that are all assumed to be known and possibly time-varying.

The above description makes it possible to introduce the recursive discrete-time Kalman-filter (giving the MMSE estimate of the system under consideration) and derive steady-state filters for noisy kinematic models (alpha-beta and alpha-beta-gamma filters). These can be then further developed and combined in adaptive estimation of maneuvering targets (e.g. interacting multiple model – IMM - approaches).

For the time being, we have embedded only a *noiseless constant velocity* kinematic state estimator while focusing on the implementation on efficient front-end filtering and data assignment strategies. Unfortunately, the more complex state estimators such as variants of the Kalman-filter or IMM state estimators [8] are computationally very intensive and will require a more advanced hardware environment for real-time MTT purposes. In order to meet these requirements we are planning to utilize a more powerful DSP (Texas C64 family) to facilitate the inclusion of more accurate state estimators.

Tracking algorithm performance

Figure 4 shows the results of running the system on two video flows that contain targets which are maneuvering and sometimes move in front of each other, effectively stress testing the tracking algorithms. The measured track states show that the system tracked the targets fairly well.

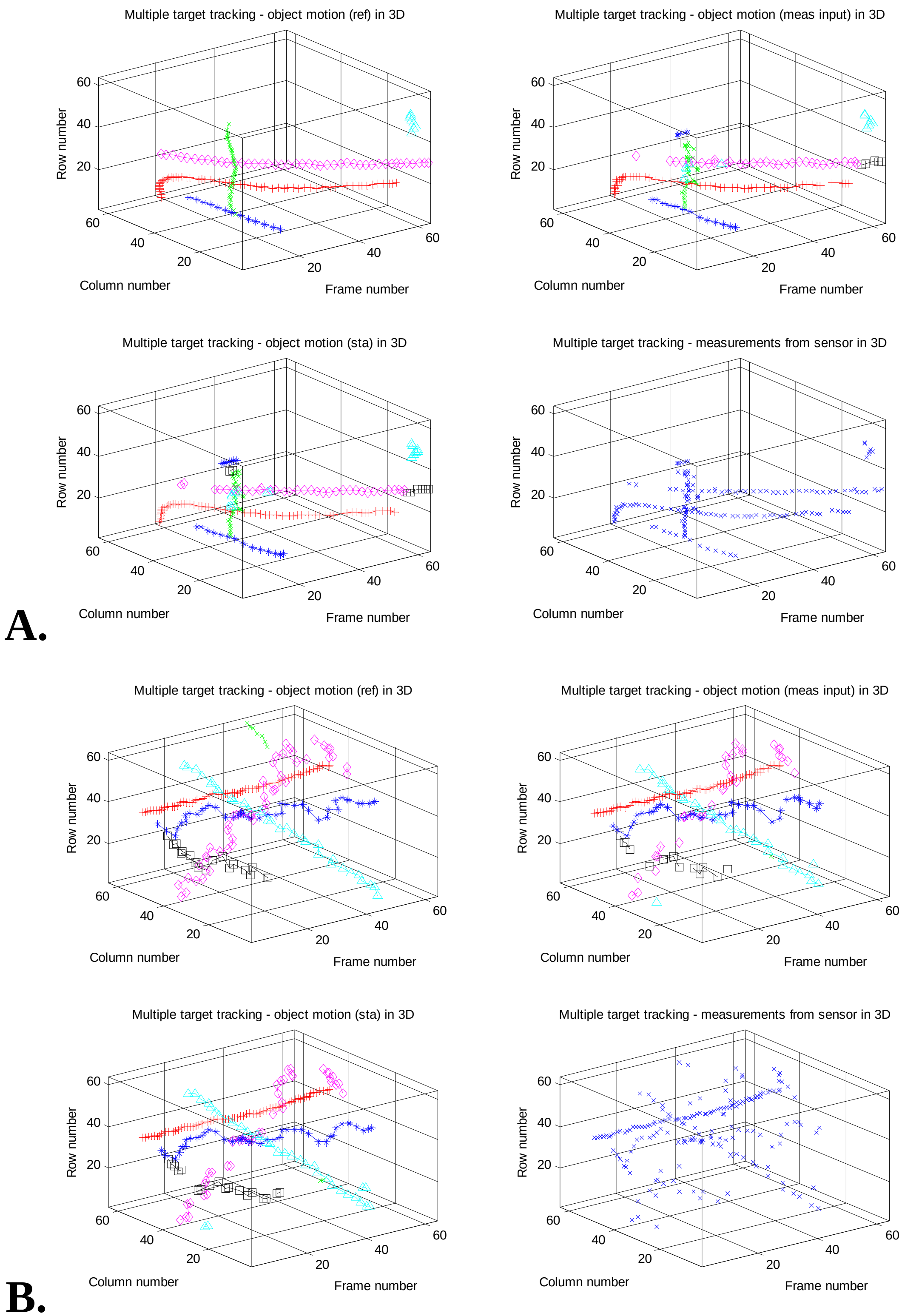

Figure 4 Tracked target positions in a sample run. The different colors signify different tracks. The reference positions ('ref', upper left plot) were marked by a human observer while the measured track states ('sta', lower left plot) are the output of the system. The system's tracking performance for different video flows: A) 'birds', B) 'cells'.

The Laser Controller

The laser controller contains the necessary electronics to translate the digital TTL signals (containing the coordinates) from the ACE-BOX into the analog voltages required to move the galvo-motors with the deflector mirrors. It also controls the ON/OFF operation of the laser itself. The galvo-mirrors are able to deflect the laser beam in ±20°-s horizontally and vertically. The whole laser apparatus can be seen on Figure 5. The laser and the deflector mirrors are affixed atop of the camera because this will keep parallax error[*] to the minimum possible.

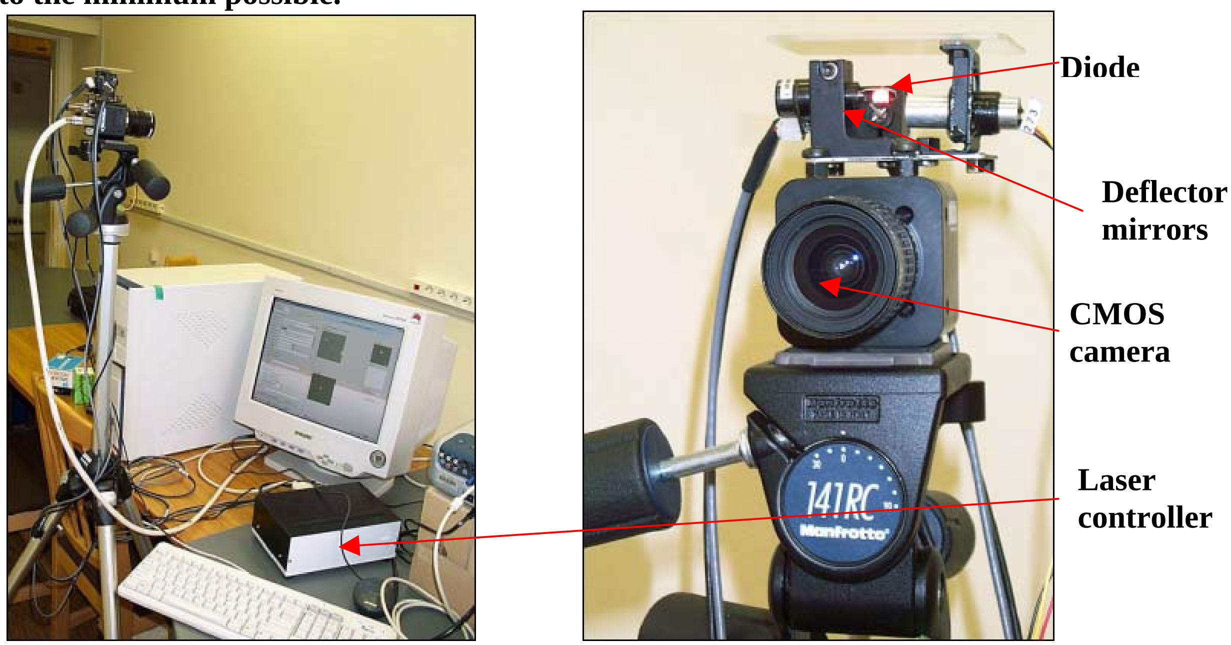

Figure 5 The laser controller and the high-speed CMOS camera (the ACE-BOX system with the ACE4k chip is located in the PC)

System speed analysis

We measured the performance of the system at the ACE16k level to determine the running time of the image processing algorithms and at the MTT algorithm level. The computational time for the multi-channel algorithms running on the ACE16k chip are approx. 4ms/channel for the image computations while the MTT algorithms take 3.5ms to run for 8 targets.

The net speed of the system (for 6 targets) is approximately 60 fps as targeted by the project specifications. Since the image processing speed does not change with the number of targets, performance of the system for more targets should be equally good.

[*] **This error arises because the optical center of the camera is not exactly aligned with that of the laser.**

References

[1] **L. O. Chua and L. Yang, "Cellular Neural Networks: Theory and Applications", *IEEE Trans. on Circ. & Syst.*, Vol. 35, pp. 1257-1290, 1988.**

[2] **L. O. Chua, and T. Roska, "The CNN Paradigm", *IEEE Trans. on Circ. & Syst.*, Vol. 40, pp.147-156, 1993.**

[3] **T. Roska and L. O. Chua, "The CNN Universal Machine", *IEEE Trans. on Circ. & Syst.*, Vol. 40, pp. 163-173, 1993.**

[4] **S. Espejo, R. Domínguez-Castro, G. Liñán, Á. Rodríguez-Vázquez, "A 64×64 CNN Universal Chip with Analog and Digital I/O", *in Proc. ICECS'98*, pp. 203-206, Lisbon 1998.**

[5] **Cs. Rekeczky, T. Roska, and A. Ushida, "CNN-based Difference-controlled Adaptive Nonlinear Image Filters", *International Journal of Circuit Theory and Applications*, Vol. 26, pp. 375-423, July-August 1998.**

[6] **G. Liñán, R. Domínguez-Castro, S. Espejo, A. Rodríguez-Vázquez, "ACE16k: A Pro-grammable Focal Plane Vision Processor with 128 x 128 Resolution", ECCTD '01 -European Conference on Circuit Theory and Design, pp.: 345-348, August 28-31,2001, Espoo, Finland**

[7] **R. Jonker R. and A. Volgenant, "A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems", *J. Computing*, Vol. 38,pp. 325-430, 1987.**

[8] **Y. Bar-Shalom W.D. Blair ed., "Multitarget-Multisensor Tracking: Applications and Advances, Vol. III", Artech House 2000.**

[9] **D.P. Bertsekas, "The Auction Algorithm: a Distributed Relaxation Method for the Assignment Problem", *Annals of Operation Research: special issue on parallel optimization*, Vol. 14, pp. 105-123, 1988.**

[10] **J. Munkres, "Algorithms for the Assignment and Transportation Problems", *J. Soc. Indust. Applied Math*, Vol. 5, No. 1, pp. 32-38, 1957.**